COMPUTATIONAL MODELING FOR CROSS-LINKING & MASS SPECTROMETRY

Chris Bailey-Kellogg*[1], Alan M. Friedman*[2], and Aly Azeem Khan[1] (aakhan@purdue.edu)

[1]Purdue University Department of Computer Sciences, 1398 Computer Science Building, West Lafayette, IN, 47907.
[2]Purdue University Department of Biological Sciences, 1392 Lilly Hall of Life Sciences, West Lafayette, IN 47907.

Proteins are directly responsible for the control of a wide array of functions in cells. Understanding a protein's function requires an understanding of its three-dimensional structure. Current analytical methodologies for determining protein structure make use of NMR and x-ray crystallography. Unfortunately, these same methodologies are demanding in time and in sample quantity and preparation. Purely computational methods, such as *threading* provide a set of possible structures for a given protein sequence [Lathrop & Smith, 1996]. Unfortunately, despite being less demanding in time, threading is limited in its ability to accurately score structural models and determine with high confidence exactly which structure is correct.

In order to address this issue of improving protein structure determination, we present here a combined computational-experimental framework for protein structure discrimination. Cross-linking & mass spectrometry [Young et al., 2000] offers a faster and easier alternative *with increased reliability* over purely computational methods for protein structure determination by incorporating experimentally derived constraints. A *cross-linker* of fixed length (e.g. such as BS^3, approx. 12Å) is introduced into a solution of a protein whose structure we wish to determine. This cross-linker then covalently bonds certain specific pairs of atoms in the protein which are within the reach of the cross-linker's length. In order to determine which cross-links were formed, proteolytic digestion (e.g. via trypsin) allows the protein to be spliced into small segments, however, keeping the cross-links intact. By using mass spectral analysis, we can determine which atoms were cross-linked with one another, and thus their relative (i.e. upper bound) distance from one another. This provides us with some spatial proximity information regarding the structure of the protein. Unfortunately, this spatial proximity information is too sparse to construct a complete protein structure. However, spatial proximity information is just the experimentally derived constraint needed to discriminate among possible protein models.

Using threading software, we generated a set of possible models for a protein. Our goal is to discriminate among the possible structural models, with the aid of experimental constraints. The key problem is assessing the protein's ability to cross-link. Measuring straight-line distance between specific atoms expected to cross-link does not provide an accurate estimate of the cross-linking distance since the straight-line distance may go through the protein. We took a more sophisticated approach by modifying a protein model with a cross-linker molecule (BS^3) and using restrained molecular dynamics techniques [Brunger et al., 1998] to estimate the cross-link distance for each possible pair of cross-links. We have also developed an alternative, completely geometric, technique for estimating the cross-link distances for a given protein structure model. We expect that these methods will allow us to create a probability function to express the cross-link likelihood for each protein structure model. With the experimental spatial proximity information, we should be able to discriminate between possible protein models by comparing which probable cross-links from a protein model fit the relative distance constraints.

This framework provides a computational-experimental approach to improve protein model discrimination through the use of experimentally derived constraints, protein distance geometry, and simulated molecular dynamics. We expect that this framework will also prove valuable for the investigation of protein-protein complex structures [Scaloni et al., 1998].

References:

A. T. Brunger et al., "Crystallography & NMR System" Acta Cryst. (1998) D54, 905-921

R. H. Lathrop, T. F. Smith, "Global optimum protein threading with gapped alignment and empirical pair potentials," J. Mol. Biol., 255:641-665, 1996.

A. Scaloni et al., "Topology of the Calmodulin-Melittin Complex," J.Mol. Biol., 277: 945- 958, 1998.

M. M. Young et al., "High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry," PNAS, 97(11):5802-5806, 2000.